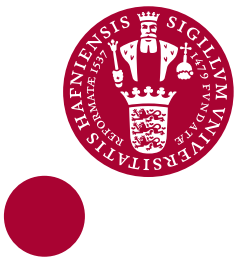

Contractile injection systems of bacteriophages and related systems

Taylor, Nicholas M.I.; van Raaij, Mark J.; Leiman, Petr G.

Published in:
Molecular Microbiology

DOI:
10.1111/mmi.13921

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Molecular Microbiology (2018) 108(1), 6–15

doi:10.1111/mmi.13921
First published online 26 February 2018

MicroReview

Contractile injection systems of bacteriophages and related systems

Nicholas M. I. Taylor [1]* Mark J. van Raaij [2] and Petr G. Leiman [3]**

[1] *Structural Biology of Molecular Machines Group, Protein Structure & Function Programme, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, Copenhagen 2200, Denmark.*

[2] *Departamento de Estructura de Macromoleculas, Centro Nacional de Biotecnologia (CSIC), Calle Darwin 3, E-28049 Madrid, Spain.*

[3] *Department of Biochemistry and Molecular Biology, 301 University Blvd, University of Texas Medical Branch, Galveston, TX 77555-0647, USA.*

Summary

Contractile tail bacteriophages, or myobacteriophages, use a sophisticated biomolecular structure to inject their genome into the bacterial host cell. This structure consists of a contractile sheath enveloping a rigid tube that is sharpened by a spike-shaped protein complex at its tip. The spike complex forms the centerpiece of a baseplate complex that terminates the sheath and the tube. The baseplate anchors the tail to the target cell membrane with the help of fibrous proteins emanating from it and triggers contraction of the sheath. The contracting sheath drives the tube with its spiky tip through the target cell membrane. Subsequently, the bacteriophage genome is injected through the tube. The structural transformation of the bacteriophage T4 baseplate upon binding to the host cell has been recently described in near-atomic detail. In this review we discuss structural elements and features of this mechanism that are likely to be conserved in all contractile injection systems (systems evolutionary and structurally related to contractile bacteriophage tails). These include the type VI secretion system (T6SS), which is used by bacteria to transfer effectors into other bacteria and into eukaryotic cells, and tailocins, a large family of contractile bacteriophage tail-like compounds that includes the *P. aeruginosa* R-type pyocins.

Accepted 1 February, 2018. For correspondence. *E-mail nicholas.taylor@cpr.ku.dk; Tel. +45 353 35337. **E-mail pgleiman@utmb.edu; Tel. +1 409 747 2078; Fax +1 409 772 5159.

Contractile injection systems

Supramolecular structure of tail sheath-like and tail tube-like proteins

A contractile injection system – a phage tail, R-type pyocin, the Type VI Secretion System (T6SS), the *Serratia* antifeeding prophage (Afp) and others – can be greatly simplified if we think of it as vertically stacked repeats of a tail tube protein hexamer surrounded by a sheath protein hexamer (Kostyuchenko *et al.*, 2003; Leiman *et al.*, 2003; Leiman *et al.*, 2004; Kostyuchenko *et al.*, 2005; Leiman *et al.*, 2010; Heymann *et al.*, 2013; Clemens *et al.*, 2015; Ge *et al.*, 2015; Kudryashev *et al.*, 2015; Taylor *et al.*, 2016; Zheng *et al.*, 2017; Wang *et al.*, 2017). The ends of this structure are terminated with a baseplate and a capping assembly (Coombs and Arisaka, 1994; Leiman and Shneider, 2012) (Fig. 1). The baseplate coordinates host recognition or another environmental signal with sheath contraction (Coombs and Arisaka, 1994; Goldberg *et al.*, 1994; Leiman *et al.*, 2004; Hu *et al.*, 2015) whereas the capping assembly prevents the tube from sliding out from the sheath during sheath contraction (Leiman *et al.*, 2004). In the phage T4 tail, all but the terminal repeats consist of the same pair of proteins: gene product (gp) 19 forms the tube and gp18 forms the sheath (Coombs and Arisaka, 1994; Leiman *et al.*, 2010). The baseplate-proximal end of the tube is formed by gp48 and gp54 hexamers (Duda et al., 1986; Ishimoto *et al.*, 1988; Watts and Coombs, 1990; Taylor *et al.*, 2016) and the distal end by a gp3 hexamer (Vianelli *et al.*, 2000). Gp3 interacts with the capping protein gp15 (Leiman *et al.*, 2004), which

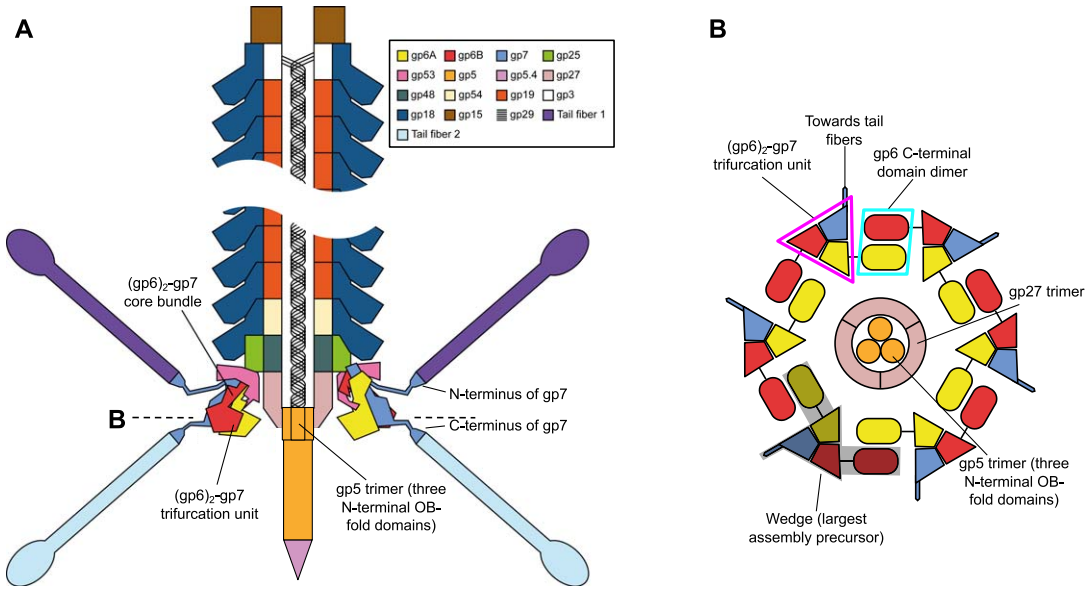

Fig. 1. Structural organization of a contractile injection system.
A. A schematic showing the composition and architecture of T4 tail. The color code is in the Inset. For simplicity, the multicomponent tail fiber network of T4 is shown as two proteins (Tail fiber 1 and 2) that are attached to the N- and C-terminus of gp7. Several features of the T4 tail (all of which are discussed in the text) are not universally conserved such as, for example, the presence of two tube initiator proteins (gp48 and gp54), the tube terminator gp3 and the interaction of tail fibers with the N-terminus of gp7-like protein. These functions are performed by other universally conserved proteins in other contractile injection systems (e.g., the same protein forms the body of the tube and the terminator). The universally conserved core bundle and the trifurcation units are labeled. The two non-equivalent copies of gp6 are in yellow and red, and gp7 is in blue. The dashed line indicates the position of the cross-section shown in panel B. Adapted from (Leiman and Shneider, 2012) and extended by data from (Taylor *et al.*, 2016).
B. A schematic representation of the circularization and fiber (receptor-binding protein) attachment mechanisms in a contractile injection system baseplate. The $(gp6)_2$–gp7 trifurcation unit and the gp6 C-terminal domain dimer are labeled with magenta and cyan lines respectively. The largest assembly precursor wedge complex is shown with a semitransparent gray shape. The schematic corresponds to the cross-section of the baseplate indicated in panel A with dashed lines. The protein color code is as in panel A.

terminates the sheath and the tube. Gp15 forms a hexamer that is larger than the tube hexamer but much smaller than the sheath disk (Leiman *et al.*, 2004; Fokine *et al.*, 2013).

The traditional view is that the gp48 hexamer is the first segment of the tube (Coombs and Arisaka, 1994; Leiman *et al.*, 2010). However, it could be argued that the tube starts even earlier, from the central spike complex in which the trimeric gp27 hub protein carries a tandem, tube protein-like domain that creates a nearly perfect hexameric ring (Kanamaru *et al.*, 2002) onto which the hexamer of gp48 is built in the assembled baseplate (Taylor *et al.*, 2016). Gp48 is surrounded by gp25, which likely functions as the sheath assembly initiator (Kudryashev *et al.*, 2015; Taylor *et al.*, 2016). All other tube proteins – gp54, gp19 and gp3 – interact with the tail sheath protein gp18 using structural motifs that are similar to the gp25–gp48 interface (Taylor *et al.*, 2016). The length of the tube is determined by a tape measure protein gp29 that also participates in the assembly of the baseplate (Abuladze et al., 1994; Leiman *et al.*, 2010). In the assembled tail, gp29 is thought to occupy the central channel of the tube in an extended state, although no high-resolution data confirming this hypothesis is available. Unlike other contractile injection systems, the T6SS does not contain a tape measure protein and the length of a T6SS organelle is often limited by the size of the cell (Basler et al., 2012; Vettiger *et al.*, 2017). However, a newly discovered subtype of the T6SS appears to possess a tape measure protein that gives such T6SS organelles a uniform length (Bock *et al.*, 2017). These T6SS assemblies are related to the *Photorhabdus* Virulence Cassettes (PVCs) (Yang *et al.*, 2006) and Afps of *Serratia* (Hurst *et al.*, 2007).

Based on studies of the T6SS (Kudryashev *et al.*, 2015; Clemens *et al.*, 2015) and R-type pyocin (Ge *et al.*, 2015), the whole sheath assembly and contractile function are presumed to rely on a β-strand exchange mechanism (Remaut *et al.*, 2006), whereby gp18 donates to and accepts the N-and C-terminal arms from other gp18 molecules (Leiman and Shneider, 2012). At the baseplate-proximal end, gp25 appears to have a vacancy for accepting gp18 strands of the first sheath layer (Taylor *et al.*, 2016). At the baseplate distal end,

gp15 could donate its N- and C-terminal arms to gp18 subunits that form the last sheath layer.

The conserved part of the baseplate

More than 24 structural proteins and chaperones participate in the assembly of T4 baseplate (Kikuchi and King, 1975b,c,d; Watts and Coombs, 1989; Watts and Coombs, 1990; Watts *et al*., 1990; Coombs and Arisaka, 1994; Leiman *et al*., 2010). Conservation of chaperones is unclear. The degree of conservation of structural components decreases with the increasing radius of their location in the baseplate (Taylor *et al*., 2016). The functions of only eight structural proteins of T4 appear to be universally conserved (Taylor *et al*., 2016). Four of them – the above-mentioned baseplate centerpiece gp48 and the central spike complex proteins gp5, gp5.4 and gp27 – form a continuation of the tube and the other four – gp6, gp7, gp25 and gp53 – create the bona fide 'plate'. Homologs of the gp48 centerpiece and gp27 hub are found in non-contractile tails, where they are called Dit and Tal respectively, making the hub complex the most conserved baseplate component of all long tail-like systems (Veesler and Cambillau, 2011).

Gp6, gp7, gp25 and gp53 are part of a large assembly intermediate that has been historically called the 'wedge' (Kikuchi and King, 1975b), although the shape of this complex is more complicated than that of a simple wedge (Taylor *et al*., 2016). Earlier biochemical data and the cryo-electron microscopy (cryoEM) derived structure of the T4 baseplate show that the wedge complex contains one copy of each gp7, gp25 and gp53 and two copies of gp6 (Watts and Coombs, 1989; Taylor *et al*., 2016). The wedge and the central hub assemble independently from each other and six wedges circularize around the hub to give rise to a complete baseplate structure (Kikuchi and King, 1975d).

In T4, all three modules of the central spike are encoded by three separate genes: the hub is gene 27, the spike is gene 5 and the spike tip is gene 5.4 (Miller *et al*., 2003). Several fusions occur in other systems, most notably of the hub and the spike in T6SS VgrG (Pukatzki *et al*., 2007; Leiman *et al*., 2009) or of the spike and spike tip in phage P2 gpV and R-type pyocin PA0616 (Browning *et al*., 2012). Spike and hub proteins often carry one or several enzymatic domains that can digest the peptidoglycan (e.g., the lysozyme domain of T4 gp5 spike or the peptidoglycan hydrolase domain of A511 gp98 hub) (Nakagawa *et al*., 1985; Kanamaru *et al*., 2002; Habann *et al*., 2014). In T4 and many other systems including the T6SS, the spike tip is formed by a single chain of the spike tip gene product that belongs to the PAAR-repeat superfamily (Shneider *et al*., 2013). In P2 gpV and some other phages, the spike tip is a domain of the trimeric spike protein (Browning *et al*., 2012). The structure is nevertheless conserved as monomeric tip proteins and tip domains in trimeric spikes consist of three similar beta-hairpins that come together to give the protein a pointy shape (Browning *et al*., 2012; Taylor *et al*., 2016).

The $(gp6)_2$–gp7 heterotrimer constitutes a large fraction of the T4 baseplate wedge (Taylor *et al*., 2016). The two copies of gp6 in this assembly are not equivalent and only certain regions of gp6 show two-fold symmetry (see below). We termed the hub-proximal part of $(gp6)_2$–gp7 heterotrimer a 'core bundle' as it is compact and consists of α-helices (Taylor *et al*., 2016). Gp25 interacts with the tip of the core bundle, and gp53 clamps the bundle roughly in the middle (Fig. 1). The hub-distal part of the $(gp6)_2$–gp7 heterotrimer forms a 'trifurcation unit'. This triangular module lies almost flat in the baseplate plane, so that two of its extensions are directed tangentially and the third is radial, away from the baseplate (Fig. 1). The radial extension is formed by gp7 and it connects tail fibers to the trifurcation unit and core bundle. The tangential extensions are formed by the C-terminal domains of two gp6 molecules, which link neighboring $(gp6)_2$–gp7 units (and, by extension, the wedges) into a ring. The dimerization of the C-terminal domains of gp6 (or its ortholog) constitutes the essence of the baseplate circularization mechanism, which is likely to be common to all contraction systems. Thus, the function of the $(gp6)_2$–gp7 heterotrimer is to circularize the baseplate and to link its peripheral interaction-sensing proteins (such as e.g., tail fibers) to the hub and the sheath.

Despite being composed of two copies of gp6 and one copy of gp7, the core bundle and the trifurcation unit are nearly perfectly threefold symmetric and the structures of the corresponding parts of gp6 and gp7 are essentially identical (Taylor *et al*., 2016). This strongly suggests that gp6 and gp7 have a common ancestor, but this information is now lost from the sequence of gp7 as well as most of its orthologs. However, bioinformatic tools that employ protein structure in their analysis are able to detect this relationship as was demonstrated by us earlier (see Supporting Information Table S3 in [Taylor *et al*., 2016]). Further confirming their relationship and common origin, the order of gp6 and gp7 homolog-encoding genes is well conserved in many systems, including more distant orthologs of T4 such as the T6SS (*tssF* and *tssG*), with the gene coding for the gp6 homologue preceding the one that encodes the gp7 homologue. Additionally, in many systems the gp25 homolog-encoding gene directly precedes the gp6 homologue-encoding gene (e.g., T4 *25–6–7* orthologous genes of phage P2 (*W–I–J*), R-type pyocin (*PA0617–PA0618–PA0619*) and T6SS (*tssE–F–G*) are sequential) (Taylor *et al*., 2016).

Gp53 performs a critical function in the baseplate, and we have postulated earlier that it is a conserved wedge protein (Taylor *et al.*, 2016). However, today's bioinformatic tools cannot identify its ortholog in some contractile injection systems (e.g., it is not found in the T6SS). Gp53 and many of its not too distant orthologs contain a LysM domain (e.g., phage P2 gpX, R-type pyocin PA0627 or A511 gp102) and this domain interacts with the $(gp6)_2$–gp7 core bundle. LysM domains are widely distributed in proteins that bind to (peptido)-glycans (Buist *et al.*, 2008) but in gp53 this domain does not display any potential sugar binding interfaces (Taylor *et al.*, 2016). The abundance of LysM domains that do bind sugars likely skews the sequence alignment profiles in favor of such proteins instead of gp53-like LysM domains thus preventing the identification of distant gp53 orthologs. A recent report (Buttner *et al.*, 2016) identifies a protein that is genetically linked to the gp25–gp6–gp7 trio but does not contain a LysM domain. This protein might nevertheless be an ortholog of T4 gp53 and contain a domain with a LysM-like structure, but its sequence is now too divergent for bioinformatics to identify this similarity. This relationship must still be present in the three-dimensional structure as is found in gp6-like and gp7-like proteins (Taylor *et al.*, 2016).

In phage T4, the baseplate assembly is the first step in the tail morphogenesis pathway (Coombs and Arisaka, 1994). Polymerization of the tube starts from the fully assembled baseplate (Kikuchi and King, 1975a). The tube-baseplate complex serves as a template for assembly of the sheath in the extended state (Arisaka *et al.*, 1979). The sheath subunits are added onto the tube starting from the baseplate (Arisaka et al., 1979; Tschopp *et al.*, 1979). The morphogenesis pathway is less studied in other systems, but similar to T4 the baseplate assembly proceeds through a large precursor or wedge in phage Mu (Buttner *et al.*, 2016) and the T6SS (Taylor *et al.*, 2016), and T6SS sheath polymerizes starting from the baseplate as well (Vettiger *et al.*, 2017). Thus, the baseplate-tube-sheath assembly sequence is likely to be common to all contractile injection systems.

Alternative composition of phage mu baseplate

The composition of the T6SS baseplate wedge produced *in vitro* is in agreement with that of T4 – the stoichiometry of TssE, TssF and TssG proteins (T4 gp25, gp6 and gp7 orthologs respectively) is 1:2:1 (Taylor *et al.*, 2016). The assembled T6SS baseplate is thus expected to have the T4-like stoichiometry of 6:12:6 (Taylor *et al.*, 2016). The baseplates of phages A511 and CBA120 have a T4-like stoichiometry and similar structure (Ricardo Guerrero and Sergey Nazarov [University of Basel], unpublished observations) showing that this organization is universal.

One recent study (Buttner *et al.*, 2016) finds that the baseplate of phage Mu is different, and the stoichiometry of Mup46–47–48 proteins (T4 gp25, gp6, gp7 orthologs respectively) is 2:2:2 in the wedge and 6:6:6 in the assembled baseplate. The different stoichiometry challenges the conservation of essentially all features of contractile injection system baseplates described earlier – the existence of the $(gp6)_2$–gp7 heterotrimer and the gp6 C-terminal domain-mediated circularization mechanism (Fig. 1).

However, there are some problems with these measurements. These findings are based on size exclusion chromatography data and densitometric analysis of SDS-PAGE gels. It is prudent to point out that size exclusion chromatography cannot be used to conclusively establish the composition in this case because the difference between the 1:2:1 (116 kDa) and 2:2:2 (154 kDa) complexes is within the experimental error for a size exclusion column. The densitometric analysis, despite its precision, can be rather inaccurate because it depends on how the Coomassie stain interacts with the protein of interest, which in turn depends on the protein sequence (Tal *et al.*, 1985). The final composition of the baseplate was derived from comparison of intensities of Coomassie-stained bands of Mup46–47–48 proteins with those of the central hub proteins, which were assumed to be trimeric, but were not shown to be so experimentally in this particular case (e.g., the hub in this case could be a dimer of trimers). At present, it is difficult to reconcile these observations with all other data and structural information on T4, CBA120, A511 and T6SS, which are very consistent with each other. Furthermore, the composition and structure of the R-type pyocin baseplate, a system that is structurally very similar to phage Mu, fully agrees with the T4 baseplate (Hong Zhou [UCLA], unpublished observations). As it was pointed out above, the likely culprit of the Mu baseplate being a putative outlier is an inconsistency in Coomassie staining (most probably, the Mu47 protein simply did not stain as well as expected). Finally, we would like to point out that a number of proteins in the T4 baseplate have somewhat abnormal Coomassie staining properties and size exclusion and SDS polyacrylamide gel mobility (Coombs and Arisaka, 1994). As a result, their stoichiometry in the baseplate remained incorrect until the crystal structure was solved (e.g., the trimeric gp11 protein [Leiman *et al.*, 2000] was originally reported to be a dimer [Plishker and Berget, 1984]).

Sheath contraction triggering

It is important to understand that tail sheath contraction must occur only upon correct attachment of the

bacteriophage to the host organism because the contraction is irreversible and bacteriophages have only one shot to successfully inject their genome into the host cell (Leiman *et al*., 2004). Inadvertent activation of tail sheath contraction will preclude the bacteriophage from replicating. Likewise, the production of a contractile injection system (e.g., T6SS) is a great expenditure of energy and material for a bacterium and its activation and triggering should be tightly regulated (Brackmann *et al*., 2017).

Sheath contraction is initiated by a structural transformation of the baseplate, which switches from a 'hexagonal' (pre-host cell attachment) to 'star' (post-attachment) conformation (Coombs and Arisaka, 1994). The latter is triggered by or coordinated with binding of the proteins comprising the periphery of the baseplate to their receptors (Leiman *et al*., 2004). In bacteriophage tails, these receptors are host cell surface structures (Goldberg *et al*., 1994). In the case of T6SS, the activation entities are unknown (but see below).

The conformational transformation of the baseplate is accomplished by a reorientation of the $(gp6)_2$–gp7 heterotrimers, which includes a small rotation and a radial translation (Taylor *et al*., 2016). The result of this transformation is the release of the hub from the rest of the baseplate and rotation of gp25 into a position that is similar to the orientation of the sheath protein gp18 in the contracted tail. Thus, gp25 appears to be important for initiation of sheath polymerization and for triggering sheath contraction.

Comparison of the baseplate structure in its two endpoint conformations shows that the transformation is accomplished by rigid body movement of domains or even complete proteins (Taylor *et al*., 2016). However, a small part of gp7 undergoes a significant refolding. One of its structured loops (residues 841–862) performs a massive, jump rope-like movement by as much as ~38 Å from one side of a β-barrel domain that is positioned 'below' the trifurcation unit to its other side. This jump rope loop might function like a cocking device for the entire baseplate structure. This mechanism involves a large energy barrier, which needs to be overcome during attachment of the phage to the host cell, when several long tail fibers bind to the cell surface.

The initial pre-attachment conformation of the baseplate is a metastable state and its ability to switch to the post-attachment state is somehow 'built into' this complex structure during assembly. How this is accomplished is unclear at the moment. We assume the jump rope loop is not universally conserved as it is part of T4 gp7 that is very divergent (Taylor *et al*., 2016). However, equivalent mechanisms must exist in other systems. All these systems are required to have a metastable state where a significant energy barrier needs to be overcome to arrive to the lower energy contracted state. This transition has to be temporally regulated to prevent an untimely firing event.

Structure of the baseplate periphery

Proteins comprising the peripheral part of the baseplate are very diverse (Taylor *et al*., 2016). In bacteriophages, these include tail fibers and tailspikes, which can range in their complexity from a single gene product (e.g., bacteriophage P2 [Haggard-Ljungquist *et al*., 1992]) to convoluted multicomponent complexes (e.g., tail fibers of A511 [Habann *et al*., 2014], CBA120 [Kutter *et al*., 2011] or AR9 [Belyaeva and Azizbekyan, 1968]) that surpass even the T4 tail fiber network in complexity, and which all but dwarf the conserved part of their baseplates.

The apparent function of the fibers is to attach the particle to the host cell surface. However, they do a lot more than just binding to some moieties on the cell surface. They are required to combine their structural and sequence diversity, which stems from an almost infinite set of ligands (cell surface structures), with a task of orienting the particle on the cell surface and positioning the baseplate such that the tube possesses sufficient thrust to penetrate the cell envelope. At the moment, it is unclear how the fiber makes it possible for the particle to 'stand up' on the cell surface, and how this property is related to the binding affinity of the fiber to its surface ligand. On the other hand, the mechanism that is employed in regulating the length of the fiber is easier to understand.

A prominent theme in the organization of T4 tail fibers is the use of sequence repeats that fold into small compact domains (Cerritelli *et al*., 1996; Taylor *et al*., 2016). Sequential duplication of a single functional unit allows the fiber to attain its elongated shape. The length of the fiber can then be controlled by the number of these units. For example, T4 short and long tail fiber proteins gp12, gp34 and gp37 contain several repeating elements (so called motifs A and B) (Cerritelli *et al*., 1996). We have recently described (Taylor *et al*., 2016) how the previously published model of the 33 kDa gp12 fragment (PDB code 1H6W [van Raaij *et al*., 2001a]) could be extended to give a complete structure of the repeating domain formed by motifs B and A (residues 220–236 and 237–253 respectively). In the process of writing this paper, we noticed that the electron density for amino acids 90–131 comprising another motif B-A domain was also of sufficient quality for model building (Fig. 2). These residues contact the next ordered layer in the crystals, as was proposed after the examination of the crystals by electron microscopy (Fig. 4 in [van Raaij

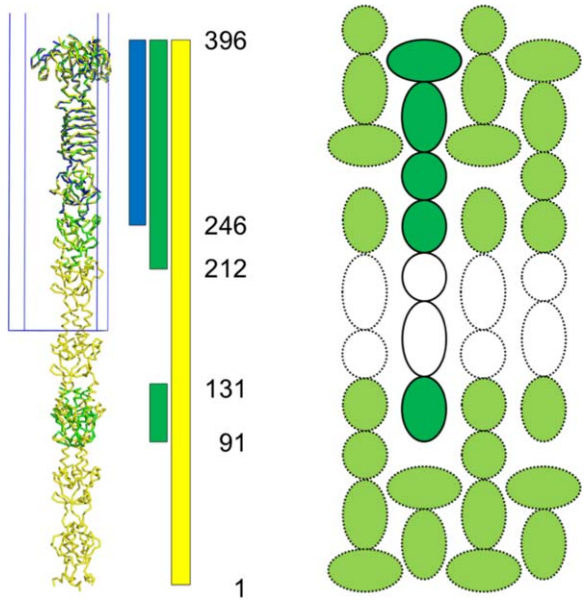

Fig. 2. Completion of PDB entry 1H6W by employing modern tools for refinement of atomic models derived from X-ray diffraction data.
A. The original structure is shown in blue and the newly refined structure in green (including the additional residues 91–131 and 212–245). The model of gp12, which was extracted from the complete baseplate structure where it is bent (PDB entry 5IV5) and computationally straightened, is shown in yellow. The unit cell of PDB entry 1H6W is shown in blue.
B. The observed packing in the crystal is shown, with ordered regions in green and regions for which no interpretable density can be observed in the crystal in white. One gp12 trimer is shown in solid lines and dark green, with surrounding trimers in light-green and dashed lines.

et al., 2001b]). We deposited the new gp12 coordinates to the Protein Data Bank under the accession number 5LYE. Recently, several B-A motifs were also resolved in the proximal long tail fiber protein gp34 (Granell *et al.*, 2017).

In contrast to bacteriophages, the T6SS does not have tail fibers. We have previously proposed that its peripheral network contains trimeric TssK proteins, which can serve as sensors that trigger the conformational change of the baseplate and subsequent sheath contraction, similar to phage tail fibers (Taylor *et al.*, 2016). Recent cryoEM reconstruction of the T6SS supports this hypothesis and places TssK at the periphery of the baseplate where it connects the baseplate to the T6SS membrane complex (Nazarov *et al.*, 2017), which is further corroborated by experimental data that show this interaction (Nguyen *et al.*, 2017). Furthermore, parts of TssK are homologous to the shoulder domain of receptor binding proteins (or 'tail fibers') of bacteriophages 1358 and p2 (Nguyen *et al.*, 2017). Thus, similar to the peripheral proteins in the T4 baseplate, TssK likely performs a sensory, conformation-triggering function in the T6SS baseplate.

Evolution of contractile injection systems

Certain structural features of T4 baseplate and their conservation across other contractile injection systems, allow us to propose a pathway by which the baseplate could have evolved starting from a simpler assembly (Fig. 3). The key component of the conserved part of the baseplate is the $(gp6)_2$–gp7 heterotrimer. Domains of gp6 and gp7 that form this heterotrimer have the same structure and thus are likely to have a common ancestor with a gp6-like fold (Fig. 3A). The N-terminus of this protein was able to form a homotrimeric core bundle whereas its C-terminal domain could dimerize and bind to some structure on the bacterial cell surface. Emergence of a separate gene encoding a gp7-like protein allowed for specialization of the receptor-binding function leading to an increased binding affinity (Fig. 3B). Finally, a duplication of the receptor-binding domain of the gp7 gene resulted in a *bona fide* tail fiber gene and allowed for fine-tuning and speeding up the evolution and adaptation process resulting in an even more specific recognition of the target cell (Fig. 3C).

Common ancestry of various domains within the tail fiber network of T4 follows from their structural similarity. For example, gp10, gp11, gp12 and gp37 contain a "flower motif" domain that is likely important for folding and trimerization (Leiman *et al.*, 2000, 2006, 2010; Bartual *et al.*, 2010; Taylor *et al.*, 2016). Furthermore, gp11 and gp12 use a homologous N-terminal domain to attach to domains 2 and 3 of gp10, which are themselves homologous to each other (Extended Data fig. 6d in [Taylor *et al.*, 2016]). This modular "mix-and-match" evolution of tail fibers appears to be a common theme for bacteriophage tail fibers, as it can be found in other bacteriophages such as CBA120 (TSP1–4) (Kutter *et al.*, 2011), phages with HK620/P22/Sf6-like tailspikes (Leiman and Molineux, 2008) and K1-specific bacteriophages (Stummeyer *et al.*, 2006).

On 'missing', 'extra', 'split' and 'fused' homologues in related systems

Certain components of the T4 tail do not (appear to) have homologs in some contractile injection systems. For example, T4 and its close relatives are unique in that they have two tube initiation proteins (gp48 and gp54), whereas there is only one tube initiator protein in most other systems (e.g., phage Mu, P2 or R-type pyocin) (Leiman and Shneider, 2012). Furthermore, bioinformatics cannot identify neither tail tube initiator nor

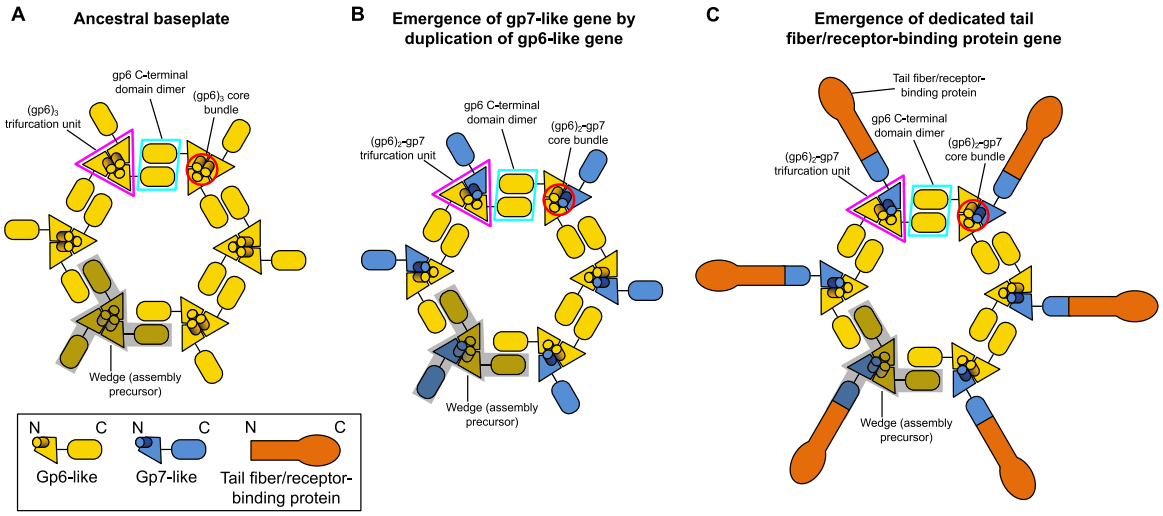

Fig. 3. Evolutionary pathway of the contractile injection system baseplate.
A. The ancestral gp6-like protein contains a trimerization module, which comprises the core bundle and the trifurcation unit, and a dimerization module that doubles as a host-interacting or sensing domain.
B. A gene duplication event results in the emergence of a gp7-like protein that is fully dedicated to host binding or some other sensory function, and the specialization of a gp6-like protein in the circularization of the baseplate.
C. The host binding/sensing domain of gp7 becomes a separate protein and the gp7-like protein turns into an adapter between the host binding/sensing protein and the baseplate. The direction of the polypeptide chain for each protein is shown in the box below panel A. In all panels, the semitransparent shape shows the assembly precursor (the 'wedge'). Note, that in some systems the tail fiber attaches to the baseplate late in the assembly process and hence it is not included into the assembly precursor.

terminator in the T6SS, suggesting that its tube protein Hcp serves as the tube initiator. Recent cryoEM structure of the T6SS complex shows a protein indistinguishable from Hcp at both ends of the tube (Nazarov *et al.*, 2017). It remains to be understood how such versatility or functional degeneracy of Hcp influences the assembly, 'power stroke' and recycling of the T6SS organelle. Binding of effectors (T6SS secreted substrates) to the central spike complex (Shneider *et al.*, 2013) or to Hcp (Whitney *et al.*, 2014) could make either or both competent for tube polymerization. In fact, tube polymerization in phages requires the tape measure protein and its chaperone in addition to the hub-like tube initiator (Coombs and Arisaka, 1994).

Whereas the aforementioned cases indicate the presence of fewer proteins that are required to create a contractile tail-like structure, the opposite is also possible. For example, the anti-feeding prophage Afp contains three proteins that show homology to the 'minimal' sheath protein (Hurst *et al.*, 2004) (two tube-proximal domains of T4 sheath protein (Aksyuk *et al.*, 2009; Leiman and Shneider, 2012) or a complete R-type pyocin sheath protein (Ge *et al.*, 2015)). It is possible that these duplicated proteins have a function in the initiation or termination of the tail sheath.

Various splitting and fusions of components can be found in different contractile injection systems (in addition to the ones already pointed out for the central spike complex). Sometimes, the function of a component is divided into separate gene products (compared to the consensus). For example, the T6SS sheath subunit is composed of two gene products (TssB and TssC), that together create a structure that is very similar to the T4 sheath subunit gp18 (Aksyuk et al., 2009; Clemens *et al.*, 2015; Kudryashev *et al.*, 2015). Conversely, the function of two proteins can be combined into a single gene product, such as is the case for *Listeria* bacteriophage A511, where protein gp102 contains a gp25- and gp53-like module (Habann et al., 2014; Taylor *et al.*, 2016).

Conclusions

The recent structure determination of the bacteriophage T4 baseplate in its pre- and post-host attachment states (Taylor *et al.*, 2016) has revealed, for the first time, the organization of a contractile injection system baseplate, as well as its transformation upon host attachment. The structure also allowed us to analyze the assembly pathway of the baseplate in quantitative terms. The baseplate is a circular, multicomponent structure with many possibilities of dividing it into six equal parts that would correspond to six putative wedges. A wedge that contains a $(gp6)_2$–gp7 heterotrimer described above (comprising the core bundle and the trifurcation unit, Fig. 1) has the lowest free energy of solvation strongly suggesting that this complex represents the true assembly precursor. The same complex is likely to be found in the

T6SS and, most probably, phage Mu baseplate wedges. Fittingly, this complex is one of the most conserved features of the baseplate, and the structure suggests its possible evolutionary origin (Fig. 3).

Many questions remain unanswered however. One of them is the quantitative role of the T4 long tail fibers in triggering the baseplate conformational switch. An answer to this question requires the knowledge of the atomic structure of the proximal part of the fiber that interacts with the baseplate (this component was not present in the T4 tube-baseplate sample). Another fundamental question is how the sheath is polymerized onto the baseplate in the extended state and why the interaction of the first layer of the sheath with the baseplate during assembly does not switch it to the post-attachment star conformation. The sheath subunit has to have a higher affinity to the post-attachment baseplate for the baseplate to be able to initiate sheath contraction by its own conformational change.

Acknowledgements

This work was funded by the grant BFU2014–53425-P (AEI/FEDER, EU) to MJvR and by the grant 310030_166383 from the Swiss National Science Foundation to PGL.

Competing financial interests

The authors declare no competing financial interests.

References

Abuladze, N.K., Gingery, M., Tsai, J., and Eiserling, F.A. (1994) Tail length determination in bacteriophage T4. *Virology* **199**: 301–310.

Aksyuk, A.A., Leiman, P.G., Kurochkina, L.P., Shneider, M.M., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* **28**: 821–829.

Arisaka, F., Tschopp, J., Van Driel, R., and Engel, J. (1979) Reassembly of the bacteriophage T4 tail from the core-baseplate and the monomeric sheath protein P18: a co-operative association process. *J Mol Biol* **132**: 369–386.

Bartual, S.G., Otero, J.M., Garcia-Doval, C., Llamas-Saiz, A.L., Kahn, R., Fox, G.C., and van Raaij, M.J. (2010) Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci U S A* **107**: 20287–20292.

Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182–186.

Belyaeva, N.N., and Azizbekyan, R.R. (1968) Fine structure of new Bacillus subtilis phage AR9 with complex morphology. *Virology* **34**: 176–179.

Bock, D., Medeiros, J.M., Tsao, H.F., Penz, T., Weiss, G.L., Aistleitner, K., *et al*. (2017) In situ architecture, function, and evolution of a contractile injection system. *Science* **357**: 713–717.

Brackmann, M., Nazarov, S., Wang, J., and Basler, M. (2017) Using force to punch holes: mechanics of contractile nanomachines. *Trends Cell Biol* **27**: 623–632.

Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., and Leiman, P.G. (2012) Phage pierces the host cell membrane with the iron-loaded spike. *Structure* **20**: 326–339.

Buist, G., Steen, A., Kok, J., and Kuipers, O.P. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* **68**: 838–847.

Buttner, C.R., Wu, Y., Maxwell, K.L., and Davidson, A.R. (2016) Baseplate assembly of phage Mu: defining the conserved core components of contractile-tailed phages and related bacterial systems. *Proc Natl Acad Sci U S A* **113**: 10174–10179.

Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F., and Steven, A.C. (1996) Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J Mol Biol* **260**: 767–780.

Clemens, D.L., Ge, P., Lee, B.Y., Horwitz, M.A., and Zhou, Z.H. (2015) Atomic structure of T6SS reveals interlaced array essential to function. *Cell* **160**: 940–951.

Coombs, D.H., and Arisaka, F. (1994) T4 tail structure and function. In: *Molecular Biology of Bacteriophage T4*. Karam, J.D. (ed.). Washington, DC: American Society for Microbiology, pp. 259–281.

Duda, R.L., Gingery, M., and Eiserling, F.A. (1986) Potential length determiner and DNA injection protein is extruded from bacteriophage T4 tail tubes in vitro. *Virology* **151**: 296–314.

Fokine, A., Zhang, Z., Kanamaru, S., Bowman, V.D., Aksyuk, A.A., Arisaka, F., *et al*. (2013) The molecular architecture of the bacteriophage T4 neck. *J Mol Biol* **425**: 1731–1744.

Ge, P., Scholl, D., Leiman, P.G., Yu, X., Miller, J.F., and Zhou, Z.H. (2015) Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat Struct Mol Biol* **22**: 377–382.

Goldberg, E., L., Grinius, and L., Letellier, (1994) Recognition, attachment, and injection. In: *Molecular Biology of Bacteriophage T4*. Karam, J.D. (ed.). Washington, DC: American Society for Microbiology, pp. 347–356.

Granell, M., Namura, M., Alvira, S., Kanamaru, S., and van Raaij, M.J. (2017) Crystal structure of the carboxy-terminal region of the bacteriophage T4 proximal long tail fiber protein Gp34. *Viruses* **9**: 168.

Habann, M., Leiman, P.G., Vandersteegen, K., Van den Bossche, A., Lavigne, R., Shneider, M.M., *et al*. (2014) Listeria phage A511, a model for the contractile tail machineries of SPO1-related bacteriophages. *Mol Microbiol* **92**: 84–99.

Haggard-Ljungquist, E., Halling, C., and Calendar, R. (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J Bacteriol* **174**: 1462–1477.

Heymann, J.B., Bartho, J.D., Rybakova, D., Venugopal, H.P., Winkler, D.C., Sen, A., *et al*. (2013) Three-

dimensional structure of the toxin-delivery particle antifeeding prophage of *Serratia entomophila*. *J Biol Chem* **288**: 25276–25284.
Hu, B., Margolin, W., Molineux, I.J., and Liu, J. (2015) Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci U S A* **112**: E4919–E4928.
Hurst, M.R., Beard, S.S., Jackson, T.A., and Jones, S.M. (2007) Isolation and characterization of the *Serratia entomophila* antifeeding prophage. *FEMS Microbiol Lett* **270**: 42–48.
Hurst, M.R., Glare, T.R., and Jackson, T.A. (2004) Cloning *Serratia entomophila* antifeeding genes–a putative defective prophage active against the grass grub *Costelytra zealandica*. *J Bacteriol* **186**: 5116–5128.
Ishimoto, L.K., Ishimoto, K.S., Cascino, A., Cipollaro, M., and Eiserling, F.A. (1988) The structure of three bacteriophage T4 genes required for tail-tube assembly. *Virology* **164**: 81–90.
Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., *et al*. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**: 553–557.
Kikuchi, Y., and King, J. (1975a) Assembly of the tail of bacteriophage T4. *J Supramol Struct* **3**: 24–38.
Kikuchi, Y., and King, J. (1975b) Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, in vivo and in vitro. *J Mol Biol* **99**: 645–672.
Kikuchi, Y., and King, J. (1975c) Genetic control of bacteriophage T4 baseplate morphogenesis. II. Mutants unable to form the central part of the baseplate. *J Mol Biol* **99**: 673–694.
Kikuchi, Y., and King, J. (1975d) Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J Mol Biol* **99**: 695–716.
Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat Struct Mol Biol* **12**: 810–813.
Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., *et al*. (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol* **10**: 688–693.
Kudryashev, M., Wang, R.Y., Brackmann, M., Scherer, S., Maier, T., Baker, D., *et al*. (2015) Structure of the type VI secretion system contractile sheath. *Cell* **160**: 952–962.
Kutter, E.M., Skutt-Kakaria, K., Blasdel, B., El-Shibiny, A., Castano, A., Bryan, D., *et al*. (2011) Characterization of a VhI-like phage specific to *Escherichia coli* O157:H7. *Virol J* **8**: 430.
Leiman, P.G., Arisaka, F., van Raaij, M.J., Kostyuchenko, V.A., Aksyuk, A.A., Kanamaru, S., and Rossmann, M.G. (2010) Morphogenesis of the T4 tail and tail fibers. *Virol J* **7**: 355.
Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., *et al*. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**: 4154–4159.
Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**: 419–429.
Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**: 2356–2370.
Leiman, P.G., Kostyuchenko, V.A., Shneider, M.M., Kurochkina, L.P., Mesyanzhinov, V.V., and Rossmann, M.G. (2000) Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. *J Mol Biol* **301**: 975–985.
Leiman, P.G., and Molineux, I.J. (2008) Evolution of a new enzyme activity from the same motif fold. *Mol Microbiol* **69**: 287–290.
Leiman, P.G., and Shneider, M.M. (2012) Contractile tail machines of bacteriophages. *Adv Exp Med Biol* **726**: 93–114.
Leiman, P.G., Shneider, M.M., Mesyanzhinov, V.V., and Rossmann, M.G. (2006) Evolution of bacteriophage tails: structure of T4 gene product 10. *J Mol Biol* **358**: 912–921.
Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003) Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**: 86–156.
Nakagawa, H., Arisaka, F., and Ishii, S. (1985) Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *J Virol* **54**: 460–466.
Nazarov, S., Schneider, J., Brackmann, P.M., Goldie, K., Stahlberg, N.H., and Basler, M. (2017) Cryo-EM reconstruction of Type VI secretion system baseplate and sheath distal end. *EMBO J* **pii**: e201797103.
Nguyen, V.S., Logger, L., Spinelli, S., Legrand, P., Huyen Pham, T.T., Nhung Trinh, T.T., *et al*. (2017) Type VI secretion TssK baseplate protein exhibits structural similarity with phage receptor-binding proteins and evolved to bind the membrane complex. *Nat Microbiol* **2**: 17103.
Plishker, M.F., and Berget, P.B. (1984) Isolation and characterization of precursors in bacteriophage T4 baseplate assembly. III. The carboxyl termini of protein P11 are required for assembly activity. *J Mol Biol* **178**: 699–709.
Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* **104**: 15508–15513.
Remaut, H., Rose, R.J., Hannan, T.J., Hultgren, S.J., Radford, S.E., Ashcroft, A.E., and Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol Cell* **22**: 831–842.
Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**: 350–353.
Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R., and Muhlenhoff, M. (2006) Evolution of bacteriophages infecting encapsulated bacteria: lessons from *Escherichia coli* K1-specific phages. *Mol Microbiol* **60**: 1123–1135.

Tal, M., Silberstein, A., and Nusser, E. (1985) Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. *J Biol Chem* **260**: 9976–9980.

Taylor, N.M., Prokhorov, N.S., Guerrero-Ferreira, R.C., Shneider, M.M., Browning, C., Goldie, K.N., *et al*. (2016) Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**: 346–352.

Tschopp, J., Arisaka, F., van Driel, R., and Engel, J. (1979) Purification, characterization and reassembly of the bacteriophage T4D tail sheath protein P18. *J Mol Biol* **128**: 247–258.

van Raaij, M.J., Schoehn, G., Burda, M.R., and Miller, S. (2001a) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J Mol Biol* **314**: 1137–1146.

van Raaij, M.J., Schoehn, G., Jaquinod, M., Ashman, K., Burda, M.R., and Miller, S. (2001b) Identification and crystallisation of a heat- and protease-stable fragment of the bacteriophage T4 short tail fibre. *Biol Chem* **382**: 1049–1055.

Veesler, D., and Cambillau, C. (2011) A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol Mol Biol Rev* **75**: 423–433 (first page of table of contents).

Vettiger, A., Winter, J., Lin, L., and Basler, M. (2017) The type VI secretion system sheath assembles at the end distal from the membrane anchor. *Nat Commun* **8**: 16088.

Vianelli, A., Wang, G.R., Gingery, M., Duda, R.L., Eiserling, F.A., and Goldberg, E.B. (2000) Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length. *J Bacteriol* **182**: 680–688.

Wang, J., Brackmann, M., Castano-Diez, D., Kudryashev, M., Goldie, K.N., Maier, T., *et al*. (2017) Cryo-EM structure of the extended type VI secretion system sheath-tube complex. *Nat Microbiol* **2**: 1507–1512.

Watts, N.R., and Coombs, D.H. (1989) Analysis of near-neighbor contacts in bacteriophage T4 wedges and hubless baseplates by using a cleavable chemical cross-linker. *J Virol* **63**: 2427–2436.

Watts, N.R., and Coombs, D.H. (1990) Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. *J Virol* **64**: 143–154.

Watts, N.R., Hainfeld, J., and Coombs, D.H. (1990) Localization of the proteins gp7, gp8 and gp10 in the bacteriophage T4 baseplate with colloidal gold:F(ab)2 and undecagold:Fab' conjugates. *J Mol Biol* **216**: 315–325.

Whitney, J.C., Beck, C.M., Goo, Y.A., Russell, A.B., Harding, B.N., De Leon, J.A., *et al*. (2014) Genetically distinct pathways guide effector export through the type VI secretion system. *Mol Microbiol* **92**: 529–542.

Yang, G., Dowling, A.J., Gerike, U., Ffrench-Constant, R.H., and Waterfield, N.R. (2006) Photorhabdus virulence cassettes confer injectable insecticidal activity against the wax moth. *J Bacteriology* **188**: 2254–2261.

Zheng, W., Wang, F., Taylor, N.M.I., Guerrero-Ferreira, R.C., Leiman, P.G., and Egelman, E.H. (2017) Refined cryo-EM structure of the T4 tail tube: exploring the lowest dose limit. *Structure* **25**: 1436–1441. e1432.